VariFunNet, an integrated multiscale modeling framework to study the effects of rare non-coding variants in Genome-Wide Association Studies: applied to Alzheimer's Disease

Qiao Liu **Biochemistry** The Graduate Center, The City University of New York New York, United States qliu2@gradcenter.cuny.edu

Chen Chen School of Computing, Informatics and Decision Systems Engineering, Arizona State University Tempe, United States chen chen@asu.edu

Lei Xie Department of Computer Science Hunter College, The City University of New York

Annie Gao Princeton High School Princeton, United States annie.gao1999@gmail.com

the Alzheimer's Disease

Hang Hang Tong School of Computing, Informatics and Decision Systems Engineering, Arizona State University Tempe, United States hanghang.tong@asu.edu

Neuroimaging Initiative ADNI investigators are at: http://adni.loni.usc.edu/wpconten t/uploads/how to apply/ADNI New York, United States Acknowledgement List lei.xie@hunter.cuny.edu

Abstract—It is a grand challenge to reveal the causal effects of DNA variants in complex phenotypes. Although statistical techniques can establish correlations between genotypes and phenotypes in Genome-Wide Association Studies (GWAS), they often fail when the variant is rare. The emerging Network-based Association Studies aim to address this shortcoming in statistical analysis, but are mainly applied to coding variations. Increasing evidences suggest that non-coding variants play critical roles in the etiology of complex diseases. However, few computational tools are available to study the effect of rare non-coding variants on phenotypes. Here we have developed a multiscale modeling variant-to-function-to-network framework VariFunNet to address these challenges. VariFunNet first predict the functional variations of molecular interactions, which result from the noncoding variants. Then we incorporate the genes associated with the functional variation into a tissue-specific gene network, and identify subnetworks that transmit the functional variation to molecular phenotypes. Finally, we quantify the functional implication of the subnetwork, and prioritize the association of the non-coding variants with the phenotype. We have applied VariFunNet to investigating the causal effect of rare non-coding variants on Alzheimer's disease (AD). Among top 21 ranked causal non-coding variants, 16 of them are directly supported by existing evidences. The remaining 5 novel variants dysregulate multiple downstream biological processes, all of which are associated with the pathology of AD. Furthermore, we propose potential new drug targets that may modulate diverse pathways responsible for AD. These findings may shed new light on discovering new biomarkers and therapies for the prevention, diagnosis, and treatment of AD. Our results suggest that multiscale modeling is a potentially powerful approach to studying causal genotype-phenotype associations.

Keywords—single nucleotide polymorphism, RNA binding, transcription factor, systems biology, network robustness, complex disease

I. INTRODUCTION

The existing methods for genotype-phenotype association are mainly based on statistical or machine learning techniques. However, the power of these methods drops significantly when sample size is small or variants are rare [1, 2]. Furthermore, these approaches work as a black box without explicitly incorporating the information of biological mechanism. As a result, it is not as straightforward to interpret the causal association of non-coding region variants with phenotypes as coding region variants. Besides, limited analysis has been performed on multi-variants that work as a system. It is believed that occurrence of genetic diseases could be caused not only by individual primary variants in coding regions but also by modest effects of multi coding and non-coding variants collectively. Small effects of non-coding variants on regulatory functions could be ignored if analysis is not performed on systematic level. Systematic analysis and multi-scale modeling of coding variants by integrating tissue-specific gene-gene interactions or protein-protein interactions is proven fruitful [3-6].

In this paper, we have developed a multiscale modeling framework to establish causal functional impact of non-coding variants on the biological network through integrating heterogeneous data from GWAS, protein-DNA/RNA interaction, protein-protein interaction, and gene expression profile, and combining statistical and machine learning

XXX-X-XXXX-XXXX-X/XX/\$XX.00 ©2017 IEEE 978-1-5090-3050-7/17/\$31.00 ©2017 IEEE

Supported by Grant Number R56AG057555 from the National Institute of Aging (NIA) of the National Institute of Health (NIH).

method, regulatory variants studies, and biological network analysis. We applied our pipeline to study the functional roles of non-coding variants and associated regulatory genes that are identified in Alzheimer Disease (AD). The appearance of GWAS and Next Generation Sequencing provide another tool for the identification of the genetic factors associated with AD. More than 20 genes have previously been proposed to be associated with AD [7]. Our method was validated by its ability to identify known risk genes, such as APOE, TOMMO40 and EGFR [7]. In addition, novel non-coding variants that may cause the dysregulation of genes such as CUX1, KHDRBS1, CNOT4, and SRSF10 were also identified as potential AD associated factors. Their down-stream regulated pathways have strong associations with AD, as supported by existing evidences. Furthermore, we propose potential new drug targets, particularly NTRK3 and CAMK2B, which may modulate diverse pathways responsible for the pathology of AD. Thus, our systematic biology approach that multi-scale models, from protein-DNA incorporates interactions in transcription level, protein-RNA interactions in translation level to gene network robustness at the system level, may shed new light on the functional roles of non-coding variants in AD in particular, and other multi-genic diseases in general.

II. MATERIALS AND METHODS

A. Data set

Data used for our method application were acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [8, 9]. It was launched in 2004 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations. ADNI is the result from efforts of many co-investigators in a broad range of academic institutions and private corporations, and subjects have been recruited from more than 50 sites across the U.S. and Canada. Approximately 200 cognitively normal older individuals were followed for 3 years, 400 people with MCI were followed for 3 years and 200 people with early AD were followed for 2 years. Our study utilized ADNI1 GWAS genetic sequencing data and AD patients' clinical data. More updated information could be found at www.loni.ucla.edu/ADNI.

B. The functional analysis of SNPs

54 SNPs were imputed to ANNOVAR [10, 11], with which SNPs position information in genes were annotated and flanking genes were indicated if it is in intergenic regions. Sequences upstream and downstream of these SNPs were retrieved from UCSC Genome Browser with hg19 as reference genome [12]. Each sequence's length is 50nt. Regulatory analysis was performed with DeepBind software [13], with which binding scores of retrieved sequences with TF and RBP were predicted. 515 TF and 120 RBP human species were included in binding study. Each SNP's two alleles differential binding scores of interactions with each regulatory protein, TF or RBP, were calculated. Then for each regulatory protein, two tailed student t-test ($H_0: \mu = 0$ and $H_a: \mu \neq 0$) was performed on 53 differential binding scores that are corresponding to 53 SNPs ($\alpha = 0.05$ and 52 degrees of freedom).

C. Analysis of gene network

The gene network of brain tissue was retrieved from Genome-scale Integrated Analysis of gene Networks (GIANT) [14]. We employed full gene network in brain, which consisted of 25689 genes, to conduct our analysis. GIANT network were based on gene-gene functional association. Gene-gene edges weights are proportional to confidence of gene-gene functional relationship and are ranged from 0 to 1.

The differential expressed genes of AD patients and normal controls came from Signaling Pathway Impact Analysis (SPIA) [15] and Common Significance Subnetworks Analysis [16] on gene expression datasets of healthy elders and AD downloaded from NCBI GEO Datasets-record of GSE5281. The dysregulated genes in these two subnetworks, including inflammation response subnetwork and calcium ion mechanism subnetwork, were involved into differentially expressed gene list. Besides, SPIA found dysregulated genes in KEGG AD pathway in hippocampus (HIP), middle temporal gyrus (MTG) and posterior cingulate cortex (PC). They also recognized perturbed genes in GABAergic synaptic pathway and glutamatergic synaptic pathway, which are two important neurotransmission system [15], in MTG, and in cytokinecytokine receptor interactions, which are key components in inflammation process [17], in PC. These dysregulated genes were also combined into the differentially expressed gene list.

To identify pathways that link the genes associated with SNPs with the differentially expressed genes, we first map these genes onto the brain tissue-specific gene network. Then network-based approach Prize-Collection Steiner Tree (PCST) algorithm was applied to identify the subnetwork that connects genes.

D. Gene-set overrepresentation analysis

5 novel putative disease associated genes' PCST subtrees, including RMBS3, CUX1, KHDRBS1, CNOT4, and SRSF10, were ranked on the top 21. Intermediate genes (excluding source and terminal nodes) in these subtrees were extracted and for gene-set overrepresentation analysis using ConsensusPathDB [18]. 13 pathway databases were integrated to this software, including KEGG, Reactome and WikiPathways. Enriched pathways passed the statistical significance threshold were listed in Supplementary Tables.

E. Calculation of Natural Connectivity

Natural connectivity is calculated as:

$$ln\left(\frac{1}{n}\sum_{i=1}^{n}e^{\lambda_{i}}\right)$$

where $\lambda_1 \ge \lambda_2 \ge ... \ge \lambda_n$ denote a non-increasing ordering of the eigenvalues of adjacency matrix of the graph.

III. RESULTS AND DISCUSSION

A. Majority of SNPs were positioned on non-coding regions

Quantitative trait association study was conducted on 620,901 single nucleotide polymorphisms (SNPs) using PLINK [19] (Figure 1) on a well-established dataset from ADNI [8, 9]. We used participants' MMSE scores on the first-visit as the phenotypes. We selected top 54 SNPs, whose



Figure 1) Schema of VariFunNet, a multi-scale modeling pipeline to study causal effect of non-coding SNPs on complex disease.

Bonferroni adjusted p-values were less than 1.0 and Benjamini & Yekutieli adjusted False Discovery Rate (FDR) values were less than 0.173. To uncover the functional roles of these 54 SNPs, we positioned them to human genome with ANNOVAR [10, 11]. This gene-based annotation was referred to Human Genome version 19 (hg19). Surprisingly, it turned out that 53 out of 54 SNPs are located in non-coding regions. This imposed a challenge to understand the functional impact of these SNPs because of very limited knowledge about the function of non-coding regions SNPs. We then proposed two ways to better understand their functions. Firstly, the flanking genes of SNPs were identified. The function of the flanking gene may shed light on the functional roles of these noncoding SNPs. Non-coding regions variants had no straightforward effect on expressed proteins sequence or structure, however they could be associated with proteins or transcripts expression level. In another word, they could vary individuals' phenotypes through their regulatory functions, especially by the cis-regulation. Secondly, transcription factor (TF) and RNA binding protein (RBP) could bind with noncoding region DNA or RNA and then either promote or inhibit proteins and transcripts expressions. TF and RBP could also collectively regulate multiple transcripts expressions or protein expressions and thus give rise to phenotype variants.

B. Identification and characterization of flanking genes

As proposed above, the flanking genes may provide a way to understand the mechanism of non-coding SNPs. ANNOVAR was used to find the flanking genes of these SNPs. In this way, totally 47 flanking genes were allocated. The most significantly associated SNP overlapped an intronic region of the most well studied AD related gene APOE. Thus, our strategy was supported by existing findings.

The second way that we proposed to study non-coding regions SNPs was to discover related TF and RBP through protein-DNA or protein-RNA interactions. TF and RBP could affect phenotype through interactions with multiple SNPs. The binding specificity of the interaction was predicted with DeepBind [13]. This software has been shown to outperform the widely used binding site prediction software, position weight matrices1 (PWMs) [13, 20]. The upstream and downstream DNA sequences of these 53 SNPs were generated using UCSC Genome Browser [12]. DeepBind was then used to generate scores for binding specificity of DNA sequences with 515 TFs. Scores differences for two different variants of each SNP were calculated. Two-tailed t-tests were performed on the score distribution of each SNP, and 27 statistical significant TFs were selected. Another 13 TFs were selected because binding scores differences of them with at least one of the 53 SNPs were higher than 4. These 40 TFs were then listed together.

C. Interpret selected genes with genetic network in system level

To better characterize the causal association of each selected SNP with AD, we integrated them into a GIANT [14]. Our method utilized PCST algorithm to find an optimal path from a gene that is associated with a SNP (source gene) to up- or down-regulated genes (terminal genes). 420 genes were used as terminal genes [15, 16, 21]. Each selected SNPassociated gene was used as source gene for the PCST. In the GIANT network, nodes denote all the identified genes in brain tissues. Each edge connects two functional related genes. The weight of each edge denotes the confidence for the functional relationship between two genes. The PCST algorithm can be used to identify a subtree that minimizes the sum of the total costs of edges included in the subtree plus the prizes of nodes not included in the subtree. In order to make the PCST work efficiently, we modified the network by deleting the edges whose weight is less than 0.3 (the edge weights ranges from 0.0 to 1.0). We mapped 1.0 minus edges weights from GIANT as the edge costs in our network in order to prioritize the edge that has high weight. We also assigned the prize of terminal node to be a fairly high value in order to contain as many terminal genes in the subtree as possible. The rest nodes have prize of 0.0. PCST could find out all the terminal genes that are related with source gene and exclude other genes from this subtree. Then the sums of total costs of edges included in the subtree plus the prizes of nodes not included in these subtrees were compared and ranked (Table 1). Lower sum value could demonstrate closer relationship of source gene with AD. Generated PCST trees were visualized with Cytoscape [22] (Figure 2).

D. Literature support of ranked genes

We manually inspected the association of top 21 ranked genes. 16 out of them in sorted gene list were found to be related to AD in the literature. For instance, Epidermal growth factor receptor (EGFR) has been reported to be potential drug target for β-amyloid peptides induced AD [23]. HFE has significant effects on iron and cholesterol metabolism, while both of these processes perform important roles in brain and are related to AD[24]. PVRL2 can mediate infection of herpes simplex. viruses (HSV), a virus found to be a risk factor for AD [25, 26]. GATA-2 activates transcription of Neuroglobin, which is implicated to reduce AD severity [27]. There is still a controversy about the effect of TOMM40 on AD, but TOMM40, as a mitochondrial transmembrane protein, is believed to be a risk gene in mitochondrial dysfunction in AD [28-30]. Other genes which are not listed here also has literature evidence for their relationship with AD.



Figure 2). Generated Prize-Collecting Steiner Tree for GATA2. Source gene GATA2 was labeled in the center. Black nodes denotes the terminal genes. Orange nodes denotes rest genes in the paths from source gene to terminal genes. PCST was visualized with Cytoscape.

Table 1: Top Ranked PCST tree sums values associated with each source gene.

Gene Symbol	PCST tree sums values	Gene Symbol	PCST tree sums values
EGFR	654.461	APOE	841.9873
HFE	720.2541	CUX1	891.3414
PVRL2	722.942	FMR1	906.7965
RBMS3	752.6028	TCF12	914.9009
GATA2	753.7089	TIA1	934.7990
NFATC1	770.7396	PTBP1	938.0070
SP1	772.2513	KHDRBS1	942.2435
EGR1	781.6129	CNOT4	956.1771
NRG1	787.6241	SRSF10	978.3997
MEF2D	800.2361	TOMM40	987.009
JUN	825.4065	POU2F1	1001.174

E. Novel SNPs that are associated with AD

Five top-21 ranked SNPs that were linked to genes RMBS3, CUX1, KHDRBS1, CNOT4, and SRSF10 did not have direct evidences about their roles in AD. However, RMBS3 is shown to directly regulate SMAD2/3 [31]. It is also reported that TGFβ-SMAD2/3 signaling inhibition is one potential therapy for AD patients [32]. Thus RMBS3 can be associated with AD through regulating TGFβ-SMAD2/3 signaling. Besides, TGFβ signaling pathway is enriched when using Gene set overrepresentation analysis (GSOA) with a q-values of 0.0205. The GSOA is also on the PCST subtree of other four discovered genes, including CUX1, KHDRBS1, CNOT4, and SRSF10, using ConsensusPathDB [18]. In the CNOT4 PCST subtree, the most significantly enriched pathways are HIF-1alpha transcription factor network, which is related to hypoxia [33], and circadian clock gene expression [34]. Increasing evidence suggests that both hypoxia and the disturbance of circadian cycle facilitate the pathogenesis of AD [35]. In the CUX1 PCST subtree, EGFR1 and FGF signaling pathway were two most significant enriched pathways. The modulation of EGFR [23] and FGF [36] have been reported to be potential therapy for the AD. The most significant pathways in KHDRBS1 PCST subtree were related to mRNA processing and splicing. In the SRSF10 PCST subtree, small nuclear ribonucleo proteins (snRNP) assembly that is related to mRNA splicing and metabolism of non-coding RNA, as well as E-cadherin adherens junction stabilization and expansion that are important for cell-cell adhension were the most significant enrichment. In summary, all five novel noncoding associations proposed in this study are supported by existing evidences.

F. Potential drug targets for the treatment of AD

We combine all pathways that are identified by the PCST into a sub-network that links genetic variants to molecular phenotypes of AD. The AD sub-network includes 1,326 genes and 6,289 directed edges. The node degree distribution in the AD sub-network follows a power-law distribution. Thus the AD sub-network has the robust-yet-fragile property of common biological network. Using the AD sub-network, it is possible for us to identify novel drug targets for the treatment of AD. Our rationale is that the potential drug target should maximally reduce the robustness of the AD subnetwork [37]. . We quantify the impact of node inhibition on the network robustness using Natural Connectivity [38]. The Natural Connectivity corresponds to an "average" eigenvalue of a graph. Natural Connectivity has clear physical and structural meaning that can be tied to several connectivity properties of networks. In particular, it explicitly characterizes the redundancy and feedback of alternative paths in the network by quantifying the weighted number of closed walks of all lengths. It has been shown that the Natural Connectivity can more accurately quantify the network robustness than other connect metrics such as node degree and betweenness etc [39]. We calculate the impact of node inhibition on the Natural Connectivity using MIoBI algorithm [39]. The top 20 genes ranked by the MIoBI score are listed in Table 2. Interestingly, three protein kinases NTRK3, CAMK2B, and PARKAR1A are ranked at top 1, 3, and 12, respectively. These protein kinases are well known druggable targets. The association of NTRK3 with neurological disorders has been revealed recently [40]. CAMK2B is a tau kinase, playing a fundamental role in neurodegeneration and memory impairment in AD [41]. The role of PRKAR1A in AD is unknown. Further studies are needed to validate this hypothesis. Another interesting potential druggable target is GPCR protein GLP1R. It has been shown that GLP1R agonist reduces AD-associated tau hyperphosphorylation[42]. Due to heavy investments in the kinase and GPCR drug discovery, it is possible for us to repurpose FDA-approved drugs for the treatment of AD. Furthermore, as AD is a multi-genic complex disease, the conventional one-drug-one-target approach is less likely to be successful. Polypharmacology may be required to develop efficient AD therapies [43]. Protein kinases and GPCRs are excellent candidates for the targeted polypharmacology [44].

Table 2. Top 20 critical nodes in the AD subnetwork determined by MIoBI. Potential kinase and GPCR drug targets are underlined.

Gene ID	Gene Symbol	MIoBI Score
4916	NTRK3	0.323
1548	CYP6A	0.585
816	CAMK2B	0.622
6426	SRSF1	0.713
56	ACRV1	0.782
7534	YWHAZ	0.797
10772	SRSF10	0.817
23369	PUM2	0.864
9444	QKI	0.874
394	ARHGAP5	0.901
6434	TRA2B	0.904
5573	PRKAR1A	0.943
6310	ATXN1	0.966
2740	<u>GLP1R</u>	1.014
2334	AFF2	1.015
6418	SET	1.019
2782	GNB1	1.044
396	ARHGDIA	1.047
9436	NCR2	1.054
860	RNX2	1.061

IV ACKNOWLEDGMENT

Data used for our method application were acquired from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [8, 9].

VREFERENCE

- R. Clarke, H. W. Ressom, A. Wang, J. Xuan, M. C. Liu, E. A. Gehan, *et al.*, "The properties of high-dimensional data spaces: implications for exploring gene and protein expression data," *Nat Rev Cancer*, vol. 8, pp. 37-49, Jan 2008.
- [2] P. J. Bickel, J. B. Brown, H. Huang, and Q. Li, "An overview of recent developments in genomics and associated statistical methods," *Philos Trans A Math Phys Eng Sci*, vol. 367, pp. 4313-37, Nov 13 2009.
- [3] H. Y. Chuang, M. Hofree, and T. Ideker, "A decade of systems biology," *Annu Rev Cell Dev Biol*, vol. 26, pp. 721-44, 2010.
- [4] I. W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, et al., "Dynamic modularity in protein interaction networks predicts breast cancer outcome," *Nat Biotechnol*, vol. 27, pp. 199-204, Feb 2009.
- [5] K. Lage, E. O. Karlberg, Z. M. Storling, P. I. Olason, A. G. Pedersen, O. Rigina, *et al.*, "A human phenome-interactome network of protein complexes implicated in genetic disorders," *Nat Biotechnol*, vol. 25, pp. 309-16, Mar 2007.
- [6] E. S. Kim, E. H. Ahn, E. Chung, and D. H. Kim, "Recent advances in nanobiotechnology and high-throughput molecular techniques for systems biomedicine," *Mol Cells*, vol. 36, pp. 477-84, Dec 2013.
- [7] L. Bertram and R. E. Tanzi, "Genome-wide association studies in Alzheimer's disease," *Hum Mol Genet*, vol. 18, pp. R137-45, Oct 15 2009.
- [8] A. J. Saykin, L. Shen, T. M. Foroud, S. G. Potkin, S. Swaminathan, S. Kim, *et al.*, "Alzheimer's Disease Neuroimaging Initiative biomarkers as quantitative phenotypes: Genetics core aims, progress, and plans," *Alzheimers Dement*, vol. 6, pp. 265-73, May 2010.
- [9] S. G. Mueller, M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, et al., "The Alzheimer's disease neuroimaging initiative," *Neuroimaging Clin N Am*, vol. 15, pp. 869-77, xi-xii, Nov 2005.
- [10] K. Wang, M. Li, and H. Hakonarson, "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data," *Nucleic Acids Res*, vol. 38, p. e164, Sep 2010.
- [11] H. Yang and K. Wang, "Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR," *Nat Protoc*, vol. 10, pp. 1556-66, Oct 2015.
- [12] W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, *et al.*, "The human genome browser at UCSC," *Genome Res*, vol. 12, pp. 996-1006, Jun 2002.
- [13] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning," *Nat Biotechnol*, vol. 33, pp. 831-8, Aug 2015.
- [14] C. S. Greene, A. Krishnan, A. K. Wong, E. Ricciotti, R. A. Zelaya, D. S. Himmelstein, *et al.*, "Understanding multicellular function and disease with human tissue-specific networks," *Nat Genet*, vol. 47, pp. 569-76, Jun 2015.
- [15] F. Chen, Q. Guan, Z. Y. Nie, and L. J. Jin, "Gene expression profile and functional analysis of Alzheimer's disease," *Am J Alzheimers Dis Other Demen*, vol. 28, pp. 693-701, Nov 2013.
- [16] W. Kong, X. Mou, N. Zhang, W. Zeng, S. Li, and Y. Yang, "The construction of common and specific significance subnetworks of Alzheimer's disease from multiple brain regions," *Biomed Res Int*, vol. 2015, p. 394260, 2015.
- [17] J. M. Rubio-Perez and J. M. Morillas-Ruiz, "A review: inflammatory process in Alzheimer's disease, role of cytokines," *ScientificWorldJournal*, vol. 2012, p. 756357, 2012.
- [18] A. Kamburov, C. Wierling, H. Lehrach, and R. Herwig, "ConsensusPathDB--a database for integrating human functional interaction networks," *Nucleic Acids Res*, vol. 37, pp. D623-8, Jan 2009.

- [19] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, *et al.*, "PLINK: a tool set for whole-genome association and population-based linkage analyses," *Am J Hum Genet*, vol. 81, pp. 559-75, Sep 2007.
- [20] G. D. Stormo, "DNA binding sites: representation and discovery," *Bioinformatics*, vol. 16, pp. 16-23, Jan 2000.
- [21] W. S. Liang, E. M. Reiman, J. Valla, T. Dunckley, T. G. Beach, A. Grover, *et al.*, "Alzheimer's disease is associated with reduced expression of energy metabolism genes in posterior cingulate neurons," *Proc Natl Acad Sci U S A*, vol. 105, pp. 4441-6, Mar 18 2008.
- [22] C. T. Lopes, M. Franz, F. Kazi, S. L. Donaldson, Q. Morris, and G. D. Bader, "Cytoscape Web: an interactive web-based network browser," *Bioinformatics*, vol. 26, pp. 2347-8, Sep 15 2010.
- [23] L. Wang, H. C. Chiang, W. Wu, B. Liang, Z. Xie, X. Yao, et al., "Epidermal growth factor receptor is a preferred target for treating amyloid-beta-induced memory loss," *Proc Natl Acad Sci U S A*, vol. 109, pp. 16743-8, Oct 9 2012.
- [24] F. Ali-Rahmani, C. L. Schengrund, and J. R. Connor, "HFE gene variants, iron, and lipids: a novel connection in Alzheimer's disease," *Front Pharmacol*, vol. 5, p. 165, 2014.
- [25] P. G. Spear, "Herpes simplex virus: receptors and ligands for cell entry," *Cell Microbiol*, vol. 6, pp. 401-10, May 2004.
- [26] R. Piacentini, G. De Chiara, D. D. Li Puma, C. Ripoli, M. E. Marcocci, E. Garaci, *et al.*, "HSV-1 and Alzheimer's disease: more than a hypothesis," *Front Pharmacol*, vol. 5, p. 97, 2014.
- [27] K. T. Tam, P. K. Chan, W. Zhang, P. P. Law, Z. Tian, G. C. Fung Chan, *et al.*, "Identification of a novel distal regulatory element of the human Neuroglobin gene by the chromosome conformation capture approach," *Nucleic Acids Res*, Sep 19 2016.
- [28] G. Jun, B. N. Vardarajan, J. Buros, C. E. Yu, M. V. Hawk, B. A. Dombroski, *et al.*, "Comprehensive search for Alzheimer disease susceptibility loci in the APOE region," *Arch Neurol*, vol. 69, pp. 1270-9, Oct 2012.
- [29] C. Cruchaga, P. Nowotny, J. S. Kauwe, P. G. Ridge, K. Mayo, S. Bertelsen, et al., "Association and expression analyses with single-nucleotide polymorphisms in TOMM40 in Alzheimer disease," *Arch Neurol*, vol. 68, pp. 1013-9, Aug 2011.
- [30] D. M. Lyall, S. E. Harris, M. E. Bastin, S. Munoz Maniega, C. Murray, M. W. Lutz, *et al.*, "Alzheimer's disease susceptibility genes APOE and TOMM40, and brain white matter integrity in the Lothian Birth Cohort 1936," *Neurobiol Aging*, vol. 35, pp. 1513 e25-33, Jun 2014.
- [31] C. S. Jayasena and M. E. Bronner, "Rbms3 functions in craniofacial development by posttranscriptionally modulating TGF-beta signaling," *J Cell Biol*, vol. 199, pp. 453-66, Oct 29 2012.
- [32] T. Town, Y. Laouar, C. Pittenger, T. Mori, C. A. Szekely, J. Tan, et

al., "Blocking TGF-beta-Smad2/3 innate immune signaling mitigates Alzheimer-like pathology," *Nat Med*, vol. 14, pp. 681-7, Jun 2008.

- [33] O. O. Ogunshola and X. Antoniou, "Contribution of hypoxia to Alzheimer's disease: is HIF-1alpha a mediator of neurodegeneration?," *Cell Mol Life Sci*, vol. 66, pp. 3555-63, Nov 2009.
- [34] N. Cermakian, E. W. Lamont, P. Boudreau, and D. B. Boivin, "Circadian clock gene expression in brain regions of Alzheimer 's disease patients and control subjects," *J Biol Rhythms*, vol. 26, pp. 160-70, Apr 2011.
- [35] X. Zhang and W. Le, "Pathological role of hypoxia in Alzheimer's disease," *Exp Neurol*, vol. 223, pp. 299-303, Jun 2010.
- [36] J. S. Li and Z. X. Yao, "Modulation of FGF receptor signaling as an intervention and potential therapy for myelin breakdown in Alzheimer's disease," *Med Hypotheses*, vol. 80, pp. 341-4, Apr 2013.
- [37] H. Kitano, "A robustness-based approach to systems-oriented drug design," *Nat Rev Drug Discov*, vol. 6, pp. 202-10, Mar 2007.
- [38] T. Hart, S. Dider, W. Han, H. Xu, Z. Zhao, and L. Xie, "Toward Repurposing Metformin as a Precision Anti-Cancer Therapy Using Structural Systems Pharmacology," *Sci Rep*, vol. 6, p. 20441, Feb 04 2016.
- [39] H. Chan, L. Akoglu, and H. Tong, "Make it or break it: Manipulating robustness in large networks," in *SIAM International Conference on Data Mining 2014, SDM 2014.* vol. 1, ed: Society for Industrial and Applied Mathematics Publications, 2014, pp. 325-333.
- [40] M. N. Braskie, O. Kohannim, N. Jahanshad, M. C. Chiang, M. Barysheva, A. W. Toga, *et al.*, "Relation between variants in the neurotrophin receptor gene, NTRK3, and white matter integrity in healthy young adults," *Neuroimage*, vol. 82, pp. 146-53, Nov 15 2013.
- [41] A. Ghosh and K. P. Giese, "Calcium/calmodulin-dependent kinase II and Alzheimer's disease," *Mol Brain*, vol. 8, p. 78, Nov 24 2015.
- [42] W. Xu, Y. Yang, G. Yuan, W. Zhu, D. Ma, and S. Hu, "Exendin-4, a glucagon-like peptide-1 receptor agonist, reduces Alzheimer disease-associated tau hyperphosphorylation in the hippocampus of rats with type 2 diabetes," *J Investig Med*, vol. 63, pp. 267-72, Feb 2015.
- [43] L. Xie, L. Xie, S. L. Kinnings, and P. E. Bourne, "Novel computational approaches to polypharmacology as a means to define responses to individual drugs," *Annu Rev Pharmacol Toxicol*, vol. 52, pp. 361-79, 2012.
- [44] Z. Zhao, L. Xie, L. Xie, and P. E. Bourne, "Delineation of Polypharmacology across the Human Structural Kinome Using a Functional Site Interaction Fingerprint Approach," *J Med Chem*, vol. 59, pp. 4326-41, May 12 2016.